UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 10/542,869 | 03/03/2006 | Morio Taneda | 9733-13 | 2415 |

20792 7590 06/16/2009

MYERS BIGEL SIBLEY & SAJOVEC
PO BOX 37428
RALEIGH, NC 27627

| EXAMINER | |
|---|---|
| ABEBE, DANIEL DEMELASH | |

| ART UNIT | PAPER NUMBER |
|---|---|
| 2626 | |

| MAIL DATE | DELIVERY MODE |
|---|---|
| 06/16/2009 | PAPER |

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

| Office Action Summary | Application No. 10/542,869 | Applicant(s) TANEDA, MORIO | |
|---|---|---|---|
| | Examiner Daniel D. Abebe | Art Unit 2626 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *20 March 2009*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *1-15* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-15* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

### *Claim Rejections - 35 USC § 103*

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

**Claims 1-3 and 13-15 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu et al. (2003/0018475) in view of Burnett et al. (2004/0133421) and further in view of Girod (6,483,532).**

As to claim 1, Basu, teaches, in "audio-visual speech detection and recognition system", a noise reduction system including an audio-visual user interface for combining visual features extracted from a digital video sequence with audio features extracted from an analog audio sequence including background noise, the system comprising:

speech sequence detection means for detecting audio signals (Par.0012-0013);

speech feature extraction and analyzing means; (Par.0038, 0042)

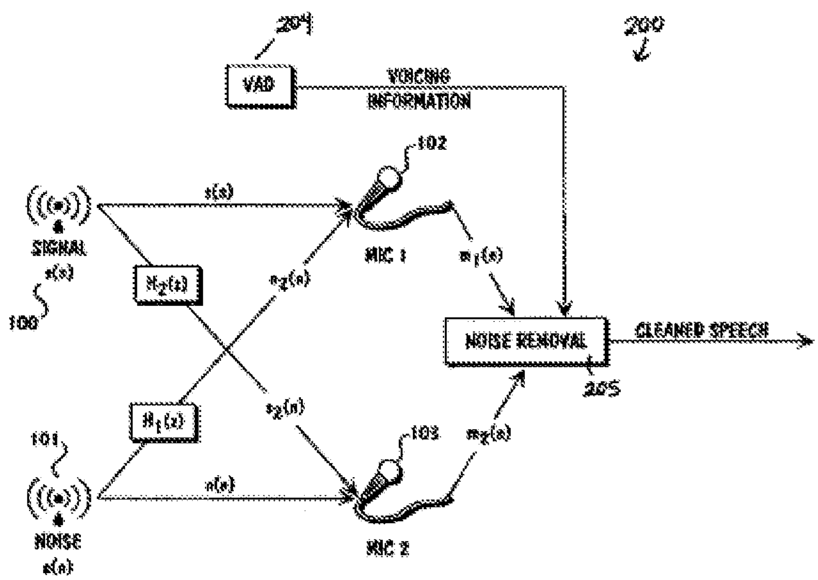
video sequence detection means for detecting said video sequence (Par.0010);

visual feature extraction and analysis means for analyzing the detected video sequence and extracting said visual features therefrom (Par.0081); and

a means to prevent background noise from being processed by the system based on the derived speech characteristics and to out put speech activity indication signal based on the combination of the speech detection and video sequence detection means (Par.0094-0097; abstract; Figs.1, 8-10; Par.0088).

it is noted that Basu doesn't explicitly teach removing noise from the speech signal.

Burnett, however, teaches a  noise suppression system (Fig.2), where the noise is remove from the speech signal according to a result of a voice activity detector wherein the voice activity detector includes motion sensors to detect the motion of the speaker (Figs.2, 6-11; Pars.0070-0076, 0082, 0089-0092).




Burnett and Basu are analogous in that they both are drawn into detecting speech using additional information to the acoustic signal, such as image and motion for the purpose of handling background noise and therefore their combination and the removing of the noise from the speech signal, in Basu teaching, is not unexpected and would be obvious for the purpose processing and removing the noise that are mixed with the speech signal of the intended speaker.

It is also noted that Basu doesn’t explicitly teach where the system comprises echo cancellation means.

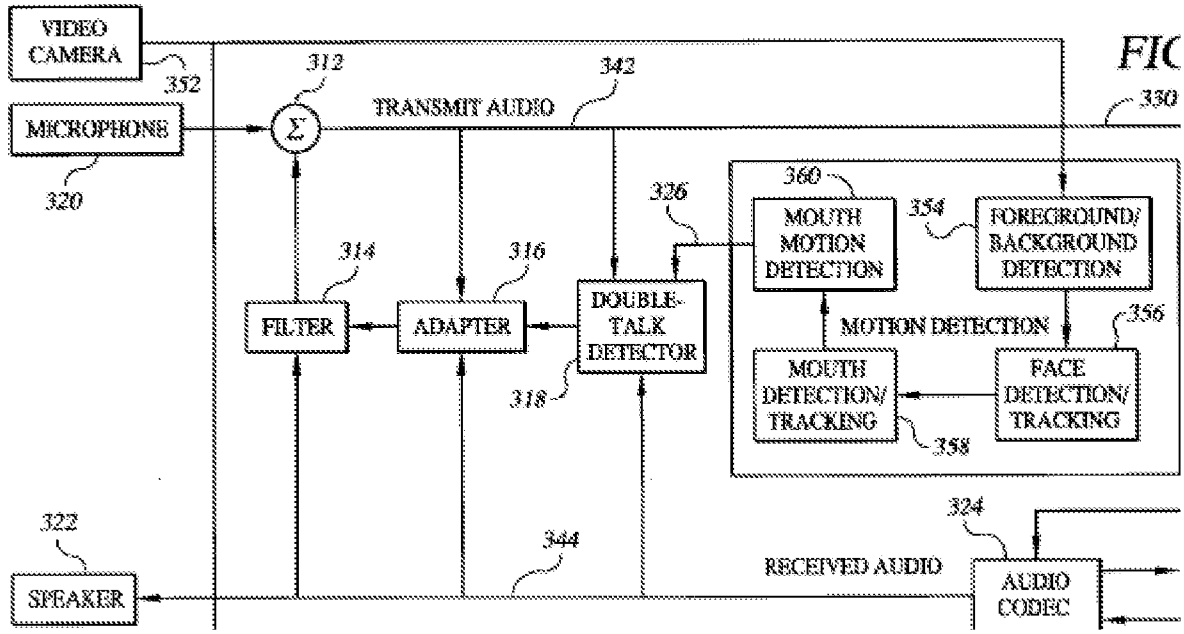
Girod, however, teaches, in a video assisted audio signal processing system, noise reduction system (Fig.3) for modifying a speech signal, including an audio-visual user interface for combining visual features extracted from a digital video sequence with audio sequence, said system comprising:

audio signal processing means (324) for processing audio signal;

video sequence detection means (354) for detecting said video sequence;

visual feature extraction and analysis means (360) for analyzing the detected video sequence and extracting said visual features therefrom; and

a multi-channel acoustic echo cancellation unit ( 312) configured to perform a near-end speaker detection (314) and double-talk detection (318) algorithm based on the audio analysis means and the visual detection means and to modify the near end speech by cancelling echo (noise) in the speech signal (abstract; Figs.1-4; Col.1, line 50-Col.2, line 38).




It would have been obvious to one of ordinary skill in the art at the time of applicant's invention to modify Basu teaching as claimed, in view of Girod, for the purpose of reliably distinguishing between speech that are meant to be processed by the system from unintended background speech including acoustic echo thereby avoiding false activation of the system.

As to claim 2, Basu teaches, enabling/disabling the microphone based on whether or not the speech energy level detected is below/above a 'given signal level' (threshold) (Par.0097, 0094, 0096).

As to claim 3, Basu teaches where the audio feature extraction and analysis means comprises an amplitude detector (Par.0039).

As to claims 13-15, Basu teaches the corresponding system for reducing noise in speech using audio features plus visual speech feature vectors as addressed above for

claim 1 in detail, and Burnett teaches removing noise from speech signal using motion detectors and acoustic detectors in telephone device and Girod teaches where the system disclosed is used in a video communication/telephony application including microphone, video camera and speaker (Figs.103; Claim 8), and the motivation for using the Basu system in video-telephony application would be obvious to one skill in the art for the purpose of reliably detecting background noise in the communication signal.

**Claims 4, 7 and 8 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu et al. (2003/0018475) in view of Burnett et al. (2004/0133421) and further in view of Wynn (5,706,394)**

As to claim 4, Basu, teaches a method for reducing noise comprising the steps of:

Converting analog speech to digital;

acoustic feature extraction process by Fourier transforming the magnitudes of discrete of samples of speech data; (Par.0038-0039, 0042); and

detecting speech in an audio signals by analyzing visual features extracted from video sequence associated with the audio sequence including current position of face, lip or facial expression of the speaker; and

preventing background noise from being processed by the system based on the derived speech characteristics and to out put speech activity indication signal on the combination of the audio processing and video sequence detection means (Par.0094-0097; abstract; Figs.1, 8-10; Par.0088).

Basu doesn't explicitly teach removing noise from the speech signal.

Burnett, however, teaches a noise suppression system, where the noise is remove from the speech signal according to a result of a voice activity detector wherein the voice activity detector includes motion sensors to detect the motion of the speaker (Figs.2, 6-11; Pars.0070-0076, 0082, 0089-0092).

Burnett and Basu are analogous in that they both are drawn into detecting speech using additional information to the acoustic signal, such as image and motion for the purpose of handeling background noise and therefore their combination and the removing of the noise from the speech signal, in Basu teaching, would be obvious for the purpose processing and removing the noise that are mixed with the speech signal that are spoken by the intended speaker.

Burnett doesn't explicitly teach the claimed process for removing noise.

Wynn teaches a method for reducing noise in speech, comprising:

Estimating a noise power density spectrum of background noise based on a voice activity detector, inherently the voice representing the user's voice;

Subtracting the estimated power noise from the speech signal;

Inverse transforming the signal into time domain where the noise subtracted speech signal could be input to speech recognizer (abstract; Col.1, lines 31-35; Col.8, line 65-Col.9, line 11; Col.16, lines14-20). It would have been obvious to one of ordinary skill in the art at the times of applicant to modify Basu system in view of Wynn for the purpose of efficiently removing background noise from the speech signal.

**As to claim 7, Basu and Burnett teach wherein said visual speech characteristics are based on detecting, face, opening of a mouth of the speaker, detecting the lips of the speaker or detecting other phonetic characteristics associated with position and movement of the lips (Par.0043-0046, Figs.2-4----Figs 3-10).**

As to claim 8, Basu teaches detecting the voice of the speaker by analyzing visual features extracted from video sequences associated with the speech where the visual features include mouth movement, face, the lips of the speaker or detecting other phonetic characteristics associated with position and movement of the lips (Par.0043-0046, Figs.2-4).

**Claims 5, 6, 9 and 10-12 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu et al. (2003/0018475) in view of Burnett et al. (2004/0133421), Wynn (5,706,394) and Girod (6,483,532).**

As to claim 5, Basu teaches where acoustic-phonetic (visual speech feature characters) are derived by an algorithm for extracting the visual feature from video sequence associated with audio sequence including movement and position of lip of facial expression in an image signal (Par.0081). Burnett teaches removing noise from speech signal using motion sensors to detect the speakers movement. the step of acoustic echo cancellation as claimed is not taught by Basu, however Girod as addressed above for claim 1, teaches a near end acoustic echo signal detection cancelling process by utilizing the combination of video detection means and audio

processing means. the motivation for combining the two teachings is same as provided in claim 1.

As to claim 6, Girod teaches where the acoustic echo cancellation process includes a double talk detection procedure (Fig.3).

As to claim 9, Wynn teaches where the noise suppressing method comprises comparing the spectrum of, inherently delayed, audio input with a voice activity estimate (threshold, TH) obtained by amplitude detection of a filtered discrete signal spectrum to provide an estimate for a frequency spectrum corresponding to a signal which represents a voice of said speaker as well as an estimate for the noise power density spectrum of the statistically distributed background noise (Fig.13; Col.14, lines14-30; Col.15, lines 2-20).

As to claims 10 and 12, Basu teaches a speech present estimation means and an event detection means where the event detection means comprises the audio feature vectors, A, extracted from audio signal and visual speech feature vectors, V, extracted from visual sequences and which are representative visual-speech and the detection is made on the combinations of the two sets of feature vectors, i.e, the audio plus the visual-speech features (Par.0042, 0080).

As to claim 12, Basu teaches where speech activity estimate features and visual-speech activity estimate features are combined/added to form a single audio visual-speech feature vector and correlated to audio visual-speech probabilities to make the detection decision (Par.103-104, 107)

Basu also teaches detecting speech using energy threshold as discussed above. Basu however, doesn't explicitly teach where speech/noise estimate is updated as claimed. Wynn teaches where the speech activity threshold is updated for every frame according to spectrally estimated noise in the speech signal (Fig.13) and this process would have been obvious in Basu system for the purpose of adjusting the energy threshold in accordance to the level of the present background noise as well as for effectively cancelling background noise in the speech signal.

As to claim 11, adjusting the frequency band of the filtered signal is inherent in Wynn teaching (Col.15, lines 2-20).

***Response to Arguments***

Applicant's arguments with respect to the claims have been considered but are moot in view of the new ground(s) of rejection.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Daniel D. Abebe whose telephone number is 571-272-7615. The examiner can normally be reached on monday-friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Hudspeth can be reached on 571-272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Daniel D Abebe/
Primary Examiner, Art Unit 2626